Title: Misinformation and DeepFake Detection on Social Media

Name: Brian Among’a
Institution: Rutgers University-Camden
Date: 02/05/2025

Introduction & Problem Statement

- ➔ Misinformation is false or untrue information, which includes rumors, insults, and pranks.
- ➔ Deepfake technology entails use of AI techniques, specifically (ML) to fabricate and manipulate audios, videos and images to deceive the target audience.
- ➔ While rapid advancements in social media has improved access to information, it has equally promoted the creation and dissemination of malicious and misleading information.
- ➔ As fake news detectors get better, fake news generators device new ways of evading the filters.

Motivation & Significance

- Misinformation raises political tension by influencing elections, public health and social stability.
- Manipulated images and videos promote hatred by inciting and demeaning others.
- Deepfake technologies encourage fraud and reputation damage that can affect individuals and companies.
- Detection helps platforms to reduce the spread of misinformation and fake news.

Research Questions & Objectives

- How effective is AI in detecting misinformation and deepfakes?
- What are the most effective machine learning techniques?
- How can real-time social media monitoring be improved?
- How can red-teaming be used to improve the existing filters and detectors?

Objectives:

- Develop misinformation and deepfake detection model
- Train the model using social media datasets
- Improve the model’s performance through red-teaming.

Data & Methodology

Data:

- FakeNewsNet - https://github.com/KaiDMML/FakeNewsNet
- Liar Dataset - https://www.cs.ucsb.edu/~william/data/liar_dataset.zip
- DFDC Dataset - https://www.kaggle.com/c/deepfake-detection-challenge/data
- FaceForensics - https://github.com/ondyari/FaceForensics

Methodology:

- NLP for text based misinformation detection
- Computer vision for deepfakes
- Ensemble learning for enhanced accuracy
- Tools: Pytorch, tensorflow, Hugging Face etc.

Possible Challenges

- Misinformation and deepfakes are increasingly becoming more realistic and harder to detect.
- High computational cost for real time detection
- Biases in data may affect detection accuracy

  Solutions
- Regular updates on deepfake datasets
- Use of Explainable AI to improve transparency
- Model optimization for efficiency in real-life datasets.

Project Expectations

- An effective AI-based misinformation and deepfake detection model
- Insights into the effectiveness of AI in addressing misinformation and deepfakes on social media.
- Applications of the model in fact-checking and social media regulation.